R E P O R T R E S U M E S

ED 010 983 RE 000 020

CLOZE READABILITY PROCEDURE.

BY- BORMUTH, JOHN R.
CALIFORNIA UNIV., LOS ANGELES

REPORT NUMBER CSEIP-OR-1 PUB DATE FEB 67
EDRS PRICE MF-$0.09 HC-$1.16 29P.

DESCRIPTORS- *CLOZE PROCEDURE, *READING COMPREHENSION, READABILITY, *PROGRAM EVALUATION, *MEASUREMENT TECHNIQUES, MEASUREMENT INSTRUMENTS, VALIDITY RESEARCH, PROGRAMED MATERIALS, LOS ANGELES

THE CLOZE READABILITY PROCEDURE FOR EVALUATING THE COMPREHENSION DIFFICULTY OF WRITTEN INSTRUCTIONAL MATERIALS WAS STUDIED. RESEARCH BEARING ON THE VALIDITY, THE FORMAL CHARACTERISTICS, AND THE APPLICATION OF THE CLOZE READABILITY PROCEDURE ARE REVIEWED. THE RESEARCH ON VALIDITY IS ARRANGED UNDER THE FOLLOWING TOPICS--MEASUREMENT OF POST-READING KNOWLEDGE, MEASUREMENT OF KNOWLEDGE GAIN, MEASUREMENT OF PASSAGE DIFFICULTY, AND CLOZE TEST RELIABILITY. RESEARCH ON APPLICATION INCLUDES--DESIGNING THE TESTING PROCEDURE, MULTIPLE SAMPLING PROBLEMS, DESIGNS, PROBLEMS, DELETION PROCEDURE, TEST ADMINISTRATION, SCORING PROCEDURE, INTERPRETATION OF SCORES, CRITERION SCORE, AND REPORTING PASSAGE PROCEDURE SEEMED TO RESULT IN VALID MEASUREMENTS OF THE COMPREHENSION DIFFICULTY OF WRITTEN INSTRUCTIONAL MATERIALS. ITS ADVANTAGES ARE THAT (1) TEST ITEMS ARE EASILY MADE, (2) IRRELEVANT SOURCES OF VARIANCE ARE NOT INJECTED INTO THE MEASUREMENT OF DIFFICULTY, (3) RESULTS ARE MORE VALID THAN PRESENTLY AVAILABLE READABILITY FORMULAS, AND (4) IT IS APPLICABLE TO A WIDE RANGE OF EVALUATION TASKS. A BIBLIOGRAPHY IS INCLUDED. (RH)

ED010983

RE 000 020

CLOZE READABILITY PROCEDURE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Center FOR THE
Study of
Evaluation
OF INSTRUCTIONAL
PROGRAMS

CSEIP Occasional Report No. 1, Feb. 1967
University of California, Los Angeles

RE 000 020

CLOZE READABILITY PROCEDURE

John R. Bormuth

CSEIP Occasional Report No. 1, Feb. 1967
University of California, Los Angeles

CLOZE READABILITY PROCEDURE

John R. Bormuth

The technology for evaluating the comprehension difficulty of written instructional materials has both instructional and economic importance. It is commonly conceded that materials should be at least minimally understandable to students, since much of what they learn is presented to them in the form of written, verbal materials. When the materials are too difficult, students fail to learn their contents. The result? The school's objectives are aborted, irreplaceable teacher and pupil time is lost, and education funds are wasted.

Cloze Readability Procedure

The purpose of this paper is to examine the cloze readability procedure, a technique that has been developed recently for use in evaluating the difficulty of instructional materials. The research bearing on the validity, on the formal characteristics, and on the applications of the cloze readability procedure will be discussed.

Cloze tests can be made in a variety of ways, but when they are used to measure the comprehension difficulties of text materials, investigators almost invariably use a specific set of procedures called the cloze readability procedure. Cloze readability tests are constructed by deleting every fifth word from a passage. The deleted words are replaced by underlined blank spaces of a uniform length, and the tests are mimeographed.

Cloze readability tests are given to subjects who have never read the passage. The subjects are instructed to fill in each blank with the word they think was deleted to form that blank. A response is scored correct when it exactly matches the word deleted. The difficulty of a passage is the mean of the subjects' percentage scores on the test.

The difficulty of every word, phrase, clause, or sentence in the passage can also be determined by using five forms of a cloze test over the passage. To make the first form words 1, 6, 11, etc., are deleted; words 2, 7, 12, etc. are deleted to make the second form. This process continues until all five forms have been constructed and each word in the passage appears as a cloze item in exactly one test form. The proportion of subjects writing the correct word in a blank is used as a measure of the difficulty of the word deleted. The difficulties of the words within a phrase, sentence, or passage are averaged to determine the difficulties of those units.

Other Evaluation Methods

Readability Formulas. Perhaps one of the chief reasons why instructional materials are not routinely evaluated to determine whether they have a suitable level of difficulty is that there has been no technique that is at once convenient, economical, and valid. Readability formulas are convenient, inexpensive, and require only unskilled clerical assistance to use, but the formulas presently available have validities that range from .5 to only about .7. Moreover, the equations take into account only a limited range

of linguistic variables and the variables that are taken into account are, by today's standards, crude. Recent research by Coleman (1966a) and Bormuth (1966a) shows that readability formulas having high validities can be developed, but the research that will obtain these formulas is still in progress.

Direct Testing. Using conventional comprehension tests to test materials directly on students seems more valid than using readability formulas, but it is also expensive and unreliable. Because the test items themselves represent a reading task for the student, it is uncertain whether it is the difficulty of the passage or the difficulty of the items that is measured by this procedure.

Programming. Instructional programming might be said to be a third method of determining the difficulty of materials. As it is currently carried out, programming is an expensive process. Furthermore, programming techniques employ test items similar to those used in conventional comprehension tests, and, in consequence, the criticisms leveled at the use of conventional comprehension tests hold also for programming.

Validity of Cloze Readability Tests

If cloze readability tests are to be used as a measure of the comprehension difficulty of written instructional materials, evidence is required showing that the tests measure the reading comprehension abilities of students. Further, it must be shown that the difficulties of cloze tests correspond to the difficulties of other tests used to measure the difficulty subjects have in understanding materials.

Criteria of Validity

Two Concepts of Comprehension. It is necessary to analyze the concept of comprehension further, since there is a fundamental disagreement about which of two measurement operations best represents the concept of comprehension ability. Traditionally, the comprehension ability of a person is measured by having him read a passage and then testing his knowledge of the content of the passage. Scores derived in this manner, however, measure both the person's knowledge acquired as a result of reading the passage and the knowledge he possessed before he read the passage. Comprehension measured in this way will be referred to as post-reading knowledge. On the other hand, many experts contend that comprehension ability is a set of generalized skills enabling the person to acquire knowledge from materials. Reasoning from this point of view leads to the claim that comprehension ability is best represented by a score obtained by finding the difference between scores on a test administered before and after the passage is read. Comprehension measured in this way will be referred to as knowledge gain.

Value Placed on Both Concepts. Both conceptualizations of comprehension are relevant to the evaluation of instructional materials. Of course it is highly desirable to select materials from which students acquire much new knowledge. But previously acquired knowledge is deliberately included in materials in order to provide the repetition essential for retention and in order to state the relationships between knowledge previously acquired and the knowledge being presented for the first time. Hence, a measure

used to assess the comprehension difficulty of materials should, ideally, be capable of measuring comprehension in either or both of these ways, since both represent desirable characteristics of materials.

Validity Research

Measurement of Post-Reading Knowledge. Nearly all the validity research on cloze readability tests has concentrated on demonstrating their validities as measures of post-reading knowledge. It seems that only one study approached this problem experimentally. Bormuth (1962) made a cloze and multiple choice test over each of nine passages, in which each passage was written so that it varied systematically in subject matter and language complexity. Both sets of tests were given to subjects in grades 4, 5, and 6. Each of the main effects and the interaction between language complexity and subject matter produced significant and roughly proportionate effects on the cloze readability and multiple choice scores.

A large number of studies have reported correlations between cloze readability test scores and scores on tests of the type to which the label comprehension is conventionally applied. The first studies discussed used comprehension tests made from the same passages as the cloze tests. Taylor (1956), using Air Force trainees as subjects, found a correlation of .76; Jenkinson (1957), using high school students, found a correlation of .82; Bormuth (1962), using elementary school pupils, found correlations ranging from .73 to .84; and Friedman (1964), using college students, gave comprehension tests consisting of 8 to 12 items each and obtained

Table 1

Correlations Between Cloze Readability Tests and Standardized Tests of Reading Achievement

Study	Subjects	Tests	Correlations
Jenkinson (1957)	High School	Cooperative Reading C2	
		Vocabulary	.78
		Level of Comprehension	.73
Rankin (1957)	College	Diagnostic Survey	
		Story Comprehension	.29
		Vocabulary	.68
		Paragraph	.60
Fletcher (1959)	College	Cooperative Reading C2	
		Vocabulary	.63
		Level of Comprehension	.55
		Speed of Comprehension	.57
		Dvorak-Van Wagenen	
		Rate of Comprehension	.59
Hafner (1963)	College	Michigan Vocabulary Profile	.56
Ruddell (1963)	Elementary	Stanford Achievement	
		Paragraph Meaning	.61-.74
Weaver and Kingston (1963, 2 cloze tests)	College	Davis Reading	.25-.5
Green (1964)	College	Diagnostic Reading Survey	
		Total Comprehension	.51
Friedman (1964) (20 cloze tests)	College (Foreign Students)	Metropolitan Achievement	
		Vocabulary	.63-.85
		Total Reading	.71-.87

correlations ranging from .24 to .43. These correlations seem high in view of the fact that, where test reliabilities were reported, the validity correlations and the reliabilities were of approximately of the same magnitude.

A fairly large number of studies have reported correlations between cloze readability tests and standardized tests of reading achievement. Table 1 shows the studies and the correlations reported. It is difficult to interpret these correlations because the authors often omitted reporting on the variances and reliabilities of the tests for the subjects used in their studies. This was a prime problem in the studies using college students. College students could be expected to exhibit a curtailed distribution of individual differences which would reduce the sizes of the correlations and, when this fact is considered, the correlations shown in Table I seem reasonably high.

Two studies investigated the factor validities of cloze tests. Weaver and Kingston (1963) performed a principle component analysis on the correlations among various tests which included some classifiable as cloze readability tests and which also included a standardized test of reading comprehension. It is interesting to note that the cloze tests exhibited low correlations with the principle component with which the comprehension test had its highest correlation. Bormuth (1966b) pointed out that this study contradicted with the findings of much of the earlier research on cloze tests, that the correlations involving other tests in the battery exhibited correlation patterns that were highly unusual for them, and that the population of subjects exhibited a curtailed range of variability.

He then presented an analysis of data from an earlier study (1962) which showed that a single component accounted for nearly all the variance in a set of cloze tests and multiple choice comprehension tests.

Measurement of Knowledge Gain. There is still only a small amount of information bearing on the question of whether cloze tests are useful as measures of knowledge gain, and even this scant information is indirect. Taylor (1956) and Rankin (1957) each found that subjects who read the intact passages before taking the cloze tests made from those passages achieved higher scores than subjects who had not read the passages. On the other hand, Green (1964) found that having subjects read the passages before taking the cloze tests did not increase their cloze scores over the scores they achieved on a cloze test given them before they read the passage. Rankin (1965) challenged Green's results pointing out that Green failed to correct for the regression effects present in studies using this design.

Measurement of Passage Difficulty. A reasonably substantial amount of research has accumulated showing that cloze readability test difficulties correspond closely to the difficulties of passages as measured by other methods. Taylor (1953), the originator of the cloze procedure, found that cloze readability test difficulties ranked the passages in the same order the readability formulas ranked them. When he selected three additional passages which, when judged subjectively, ranked one way, though when analyzed by readability formulas, ranked in the reverse order, the cloze readability test difficulty rankings agreed with the subjective judgments. Sukeyori

(1957) found a correlation of .83 between the combined subjective rankings given eight passages by three judges and cloze readability test difficulties of the passages. Bormuth (1962) found a correlation of .92 between the cloze readabilities of 9 passages and the difficulties of multiple choice comprehension tests made from the same passages. In a more recent study, Bormuth (1966) used four sets of 13 passages each and found correlations ranging from .91 to .96 between the cloze readabilities and the comprehension difficulties of the passages. The correlations between the mean number of words pronounced correctly by subjects who read the passages orally and the cloze readabilities of the passages ranged from .90 to .95.

Cloze Test Reliability. When cloze readability tests are used only as measures of the relative abilities of subjects, they are probably somewhat less reliable than well made multiple choice tests containing the same number of items. For example, Bormuth (1962) found that the reliabilities of the nine, 31 item multiple choice tests used in his study exhibited reliabilities about equal to those of the nine, 50 item cloze readability tests made from the same passages. It seems likely that this may have resulted from the fact (Fletcher 1959 and Bormuth 1962) that cloze readability tests nearly always contain a number of very difficult and very easy items which are less efficient discriminators (Davis 1949) than items in the intermediate range of difficulty. However, the large number of very difficult and very easy items appearing in cloze readability tests is actually an asset, making the tests useful in testing subjects differing widely in ability. Zero scores, maximum scores, and

skewed distributions are rarely observed when cloze readability tests are carefully administered. But this range apparently has its limits. Gallant (1964) found that cloze test reliability was reduced sharply when the tests were used with first grade children.

Application of the Cloze Readability Procedure

A substantial body of research has dealt with the technical questions arising when cloze readability procedure is used to evaluate the difficulty of instructional materials. The results of this research seem to justify the application of the procedure to a range of evaluation tasks. The following discussion takes up the major problems encountered at each step and discusses the research dealing with those problems.

Designing the Testing Procedure

Cloze readability procedure may be adapted either to measuring the difficulties of short or long passages or to measuring the difficulty of a given piece of material for an individual or for a whole group. Because the number of possible testing designs are almost infinite, only three designs will be discussed to illustrate the principles and problems of designing materials evaluation studies.

Multiple Sampling Problems. When the cloze readability procedure is used to determine the difficulty of a text, the investigator often deals simultaneously with three samples. First, because it is often impractical to test materials on the whole population with whom the materials are to be used, the investigator draws a

sample of pupils to represent this population. The accuracy of his results depends, in part, on the extent to which the sample is representative of the population.

Second, the items in a cloze test represent only a sample of the items that can be made over that passage. When long texts are evaluated, it may be an inefficient use of resources to make all five of the cloze test forms over the passages studied. As a result, the investigator must sometimes deal with what is called item sampling error. The Kuder-Richardson (1937) formula 21 for calculating test reliability takes item sampling error into account (Lord 1955). The error of the mean that is due to item sampling error may be usefully estimated by Lord's (1955) formula 21. A simpler procedure is to use two or more cloze test forms over the same passage, and then calculate the variance of the form means. Subtracting the population sampling error variance from the variance of the form means gives an estimate of the item sampling error.

Third, when a lengthy text is evaluated, it is generally not practical to make a cloze test over its full extent so sample passages must be drawn from the text and the cloze tests made over just the sample passages. Hence, the investigator must consider passage sampling error. Passage sampling error can be estimated by finding the difficulty of each of the passages in the sample, calculating the variance of the passage difficulties and then subtracting the population and item sampling error variances.

Designs. An elaborate design for a text evaluation study might follow these steps: first, the sections of the text are numbered consecutively and passages drawn randomly from each chapter.

Two or more passages are drawn from each chapter so that the relative difficulties of different chapters can be compared; second, two or more forms of a cloze test are made from each passage. The tests should be nearly identical in the number of items they contain; and third, the sample of pupils is drawn randomly, or as nearly so as possible, from the population with whom the cloze tests are to be used, and each pupil is randomly assigned to take one of the cloze tests. When two or more texts are being evaluated, this design permits the investigator to use analysis of variance to determine whether if the materials differ significantly and to determine how variable each text is from chapter to chapter.

A less expensive procedure consists of using shorter passages, say 50 words in each. Two forms of a cloze test are made from each passage and the passages are formed into a single test having two forms. The tests are then given to pupils drawn randomly from the population. Thlis procedure also permits the comparison of two or more different texts, but it does not permit the comparison of chapters within a text. It is also less reliable because shorter passages were used.

The simplest problems are presented by the evaluation of short passages such as test items, picture captions, and other passages of less than about 1,000 words. All five forms of a cloze test are made from the passage and each form is given to a different randomly selected sample of pupils. Where the passage is very short, (containing fewer than about 30 items), it is doubtful that individual scores are sufficiently reliable to permit an accurate judgment of h[illegible] well a given individual understood the passage, but the results

provide an accurate estimate of how well the group as a whole understood the passage.

Problems. The first problem encountered is the decision of how many pupils, cloze test items, and sample passages should be used. Increasing the number of each reduces the error in estimating the difficulty of the materials, but by different amounts. Bormuth (1965a) found that increasing the number of items in a cloze test reduces error more rapidly than adding the same number of students, but there is presently no knowledge of the relative size of the error resulting from passage sampling. The second problem stems from the conjecture that the difficulty of a sample passage from a text may depend in some degree on whether the pupil has studied the text preceeding the passage. While this may present little problem in most content areas, it is conceivable that in areas such as science, the effect could be considerable. This would seem to indicate that some evaluation studies should be designed to accompany instruction in such a way that the pupil is tested on a passage just before he is to study the section containing that passage.

Deletion Procedure

While nearly all readability research employs tests made by deleting every fifth word, cloze tests can be made by deleting every nth word, words at random, or just the words of a given type. The only restriction is that the words deleted must be selected entirely by an objectively specifiable process, otherwise the test must be classified as a common completion test (Taylor 1953).

Cloze test users encountered the problem of discovering how many words of text had to be left between cloze items. Leaving fewer words between items makes it possible to obtain a larger number of items from a given length of text and reduces the number of test forms that have to be made in order to eliminate item sampling error. But leaving too few words between items introduces the possibility that items will exhibit statistical dependence of the sort where the probability of a subject responding correctly to an item is dependent upon whether he is able to answer adjacent items. When appreciable statistical dependence exists, test scores cannot be treated by conventional statistical methods. MacGinitie (1961) studied the problem by varying the number of words of text left intact on either side of a set of cloze items. He was unable to detect any dependence among items when four or more words of text were left between items.

Taylor (1955) pointed out that methods involving the deletion of only words belonging to certain categories had to be excluded for use in readability studies because the frequency with which such words occur in a passage may itself be a variable influencing the difficulty of the passage. There seems to have been no research dealing with some of the more technical problems in the deletion process such as the problem of what should be deleted when a numeral is encountered. For example, should 128 be treated as if it contained three words or should it be deleted as a unit? It is not even clear if a criterion can be found for deciding issues of this sort.

Test Administration

The two principle alternatives in administering a cloze test are to give it either to subjects who have not read the passage or to subjects who have first been exposed to the passage. Giving the cloze test to subjects who have not read the passage obviously economizes on time. Moreover, it might be argued that giving a cloze test to subjects after they have read the passage causes scores to be influenced by the subject's rote memorization of the passage. (Rote memory is a learning process commonly regarded as being different from comprehension.)

The results of validity studies indicate that it makes little difference which method is used. For example, Taylor (1956) found that scores on cloze tests administered after subjects had read the passages exhibited both slightly greater variances and slightly higher correlations with comprehension tests than cloze tests administered to subjects who had not read the passage. Rankin's (1957) studies showed the same results. The greater variance alone seems sufficient to account for the increased correlation. Consequently, when greater validity or reliability are desired, it is probably more economical to obtain it by increasing the number of items in the cloze test and by giving the tests to subjects who have not read the passage.

Scoring Procedure

A response can differ from the deleted word in semantic meaning, grammatical inflection, and spelling. Users of cloze readability tests nearly always score correct just those responses where the stem of the response, the uninflected form of the word, exactly

matches the word deleted. The research seems to support this practice. Taylor (1953) found that scores obtained by counting synonyms, in addition to responses exactly matching deleted words, were no better than scores obtained by counting only responses exactly matching the words deleted when the scores were used to discriminate among passage difficulties. Rankin (1957) and Ruddell (1963) found that scores obtained by counting words exactly matching and synonyms of the deleted words resulted in the scores' having slightly, but not significantly, greater variances and correlations with scores on comprehension tests.

In the past, some investigators scored responses correct when they were inflected differently from the deleted word. Bormuth (1965b) studied the correlations between comprehension test scores and several categories of cloze test scores which were obtained by counting responses classified according to whether their inflections were correct in the context of the blank and further classified them according to whether the stem of the response exactly matched, was synonomous with, or was semantically unrelated to the deleted words. All scores obtained by counting grammatically correct responses exhibited positive correlations. The correlation involving a count of exactly matching responses was .84; the one involving a count of synonyms was .64; and the one involving semantically unrelated responses was .56. All other correlations were either negative or so small as to be indistinguishable from zero. Furthermore, a multiple regression analysis indicated that scores based on a count of the responses which exactly matched the deleted words in both inflection and word stem accounted for 95 per cent of the comprehension test

variance that could be predicted from the total set of cloze test scores. It would seem, therefore, that the most economical and objective method of scoring cloze tests, the exact word method, yields the most valid results.

Most investigators score misspellings correct when the response is otherwise correct and when the misspelling does not result in the correct spelling of another word that also fits the syntactic context of the blank. No research seems to have tested the validity of this practice. Similarly, the influence of illegibly written responses has not received study.

Interpretation of Scores

The difficulty of a text should be reported in terms that make clear how appropriate the text is for a given individual or group. This may be done either by stating the proportion of the group which is able to achieve cloze readability scores at or above some criterion level of performance or by stating the level of achievement possessed by pupils who are able to attain the criterion level of performance. To do either requires that a criterion score on cloze readability tests be established as representing an acceptable level of understanding of passage.

Criterion Score. Establishing a criterion of acceptable performance on a cloze readability test presents two major problems. First, since cloze readability tests have been in use for only a short time and since they differ radically in difficulty from conventional tests, users have not yet developed a "feel" for what is acceptable performance on a cloze test. Second, the establishment

of a criterion score has traditionally been viewed as a matter to be left to personal preference or arbitrary choice rather than as a matter for rational decision based, at least in part, on empirical data.

The most direct approach to establishing a criterion score for cloze readability tests is to adopt a criterion score traditionally used and then to determine what cloze score is comparable to this criterion score. Bormuth (1966c and 1966d) adopted the 75 per cent criterion score which has a long tradition of acceptance (Thorndike 1917) and widespread use in current practice (Betts 1946 and Harris 1962). According to this criterion, a passage is said to be suitable for use in a pupil's instruction if he responds correctly to 75 per cent or more of the questions asked him about the passage. In one study, Bormuth used multiple choice tests and had the pupils read the passages silently. In the other study, using different materials and subjects, he used short answer completion tests and had the pupils read the passages and respond to the questions orally. In both studies a cloze score of about 44 per cent was found to be comparable to the 75 per cent criterion. Since the exact word method of scoring was used in both studies, this cloze criterion score is useful only for interpreting other cloze readability tests scored according to that method.

A more adequate approach to the establishment of a criterion score was demonstrated by Coleman (1966b), who set out to determine what level of passage difficulty resulted in the greatest amount of information gain on the part of students reading the passages. He measured information gain by typing the passage on a transparency

and covering the words with strips of tape. When this was projected, the student was asked to guess and write down the first word. Then, that word was exposed and the student was asked to guess the next one. Following the first run through the passage, the tape was replaced and the procedure repeated. The difference between a student's scores on the two trials was taken as a measure of information gain. Passage difficulty was determined on a matched group of subjects using cloze readability tests. Interestingly, Coleman's results seemed to show that maximum information gain occurred on passages having difficulties of close to 44 per cent, and the cloze score was found to be comparable to the traditional 75 per cent criterion. A question has been raised (MacGinitie 1966) about whether the "information gained" by the subjects in Coleman's study was unduly influenced by rote memorization. Whatever the merits of that conjecture, it seems clear that this study demonstrated how a rational approach can be made to the establishment of criterion scores.

Reporting Passage Difficulty. The simplest method of reporting difficulty scores is to report the mean difficulty of the text and the proportion of subjects whose score exceeded the criterion score. This method, however, limits the general usefulness of the results ---it is often impossible to draw the subjects in such a way that they are a representative sample of the pupils with whom the materials are to be used, so there is no way to be sure that the proportion of subjects who reached the criterion score in the sample will represent the proportion in the population. What's more, even if the sample of subjects should be representative of the population in a school

system, it is virtually certain that the sample is not representative of subjects in the total population of pupils with whom the materials are to be used. Since text readability studies are of general interest and since they are somewhat costly to conduct, it seems advisable to use a somewhat more generally useful method of reporting the difficulty of a text.

A fairly easy method is to use results where a grade placement number is given to the text. First, the subjects' scores on the cloze readability tests are correlated with their scores on a test of reading achievement. Then, using the regression prediction formula, the achievement grade placement score that corresponds to the cloze readability criterion score is calculated. Next, the grade placement score is interpreted as the average achievement of subjects who were able to attain the criterion level of performance on the cloze tests made from the text. Other schools using the same achievement test can estimate the appropriateness of the text for their pupils by determining what proportion of the pupils have achievement scores that exceed the passage grade placement reported. And, since there are many published studies of the comparability of achievement test norms, the results should be useful regardless of what achievement test a school uses.

Conclusions

The use of the cloze readability procedure seems to result in valid measurements of the comprehension difficulty of written instructional material. The correlations between cloze readability and conventional comprehension test scores are high, and none of the

research has presented convincing evidence that the processes employed in responding to cloze readability tests are in any major sense distinguishable from those employed in responding to conventional comprehension tests. Moreover, passage difficulties determined using cloze readability tests correspond closely to the passage difficulties obtained using other measures.

The cloze readability procedure has a number of advantages not shared by other available methods of determining difficulty. Unlike the conventional test items used in other methods where materials are tried out directly on students, cloze test items are easily made and do not inject irrelevant sources of variance into the measurement of difficulty. Furthermore, cloze readability procedure yields far more valid results than the readability formulas presently available. However, when the readability formulas now in development become available for general use, they will probably be almost as valid and much less costly to use than the cloze readability procedure.

Research on the technology of the cloze readability procedure seems sufficient to permit the application of this procedure to a wide range of materials evaluation tasks, but three important problems remain to be solved: first, it is not at all certain whether cloze readability tests can be used to measure knowledge gain; second, a criterion level of performance has yet to be established on a rational basis; and third, it has yet to be determined if the act of isolating a passage from its context affects the difficulty of the passage. There are also a few other problems such as the question of how to handle numerals in the word deletion rules. None of these problems, however, seriously impairs the usefulness of the cloze readability procedure in improving the quality of materials evaluation studies.

Bibliography

1. Betts, E. A., Foundations of Reading Instruction. New York: American Book Company, 1946.

2. Bormuth, J. R., "Readability: A New Approach," Reading Research Quarterly, 1 (Spring, 1966a) 79-132.

3. Bormuth, J. R., "Factor Validity of Cloze Tests as Measures of Reading Comprehension Ability," (paper read at the American Educational Research Association in February, 1966b).

4. Bormuth, J. R., "Cloze Test Readability: Criterion Reference Scores," 1966c (in press).

5. Bormuth, J. R., "Comparable Cloze and Multiple Choice Comprehension Test Scores," 1966d (in press).

6. Bormuth, J. R., "Optimum Sample Size and Cloze Test Length in Readability Measurement," Journal of Educational Measurement, 2 (June, 1965a) 111-116.

7. Bormuth, J. R., "Validities of Grammatical and Semantic Classifications of Cloze Test Scores," Proceedings of the International Reading Association, 10 (1965b) 283-286.

8. Bormuth, J. R., Cloze Tests as Measures of Readability. Unpublished doctoral dissertation, Indiana University, 1962.

9. Coleman, E. B., "Developing a Technology of Written Instruction: Some Determiners of the Complexity of Prose" Symposium on Verbal Learning Research and the Technology of Written Instruction, Columbia University, 1966a, unpublished.

10. Coleman, E. B. and G. R. Miller, "A Measure of Information Gained During the Learning of Prose," Texas Western College of the University of Texas, 1966b, unpublished.

11. Davis, F. B., Item Analysis Data Their Computation, Interpretation, and Use in Test Construction. Harvard Education Papers, No. 2, Cambridge: Harvard University, 1949.

12. Fletcher, J. E., A Study of the Relationships Between Ability to Use Context as an Aid in Reading and Other Verbal Abilities. Unpublished doctoral dissertation. University of Washington, 1959.

13. Friedman, M., The Use of the Cloze Procedure for Improving the Reading Comprehension of Foreign Students at the University of Florida. Unpublished doctoral dissertation, University of Florida, 1964.

14. Gallant, R., An Investigation of the Use of Cloze Tests as a Measure of Readability of Materials for the Primary Grades. Unpublished doctoral dissertation. Indiana University, 1964.

15. Greene, F. P., A Modified Cloze Procedure for Assessing Adult Reading Comprehension. Unpublished doctoral dissertation, The University of Michigan, 1964.

16. Hafner, L. E., "Relationships of Various Measures tc the Cloze," In Eric L. Thurston and Lawrence E. Hafner (Eds.), Thirteenth Yearbook of the National Reading Conference, Milwaukee: National Reading Conference, Inc., 1963, 135-145.

17. Harris, A. J., Effective Teaching of Reading. New York: David McKay Company, Inc., 1962.

18. Jenkinson, M. E., Selected Processes and Difficulties in Reading Comprehension. Unpublished doctoral dissertation, University of Chicago, 1957.

19. Kuder, G. F. and M. W. Richardson, "The Theory of the Estimation of Test Reliability" Psychometrika, 2 (1937) 151-160.

20. Lord, Frederic M., "Sampling Fluctuations Resulting from the Sampling of Test Items," Psychometrika, 20 (March, 1955) 1-23.

21. MacGinitie, W. H., "Contextual Constraint in English Prose Paragraphs," Journal of Psychology , 51 (January, 1961) 121-130.

22. Rankin, E. J. "Cloze Procedure-A Survey of Research," Yearbook of the South West Reading Conference, 14 (1965), 133-148.

23. Rankin, E. F., Jr., An Evaluation of the Cloze Procedure as a Technique for Measuring Reading Comprehension. Unpublished doctoral dissertation, The University of Michigan, 1957.

24. Ruddell, R. B., An Investigation of the Effect of the Similarity of Oral and Written Patterns of Language Structure on Reading Comprehension. Unpublished doctoral dissertation, Indiana University, 1963.

25. Sukeyori, S., "A Study of Readability Measurement -- Application of Cloze Procedure to Japanese Language," (English Abstract), Japanese Journal of Psychology 28 (August, 1957), 135.

26. Taylor, W. L., " 'Cloze Procedure': A New Tool for Measuring Readability," Journalism Quarterly, (Fall, 1953) 115-432.

27. Taylor, W. L., "Recent Developments in the Use of 'Cloze Procedure'," Journalism Quarterly 33 (Winter, 1956) 42-48+.

28. Thorndike, E. L., "Reading and reasoning: A study of mistakes in paragraph reading," Journal of Educational Psychology 8 (1917) 323-332.

29. Weaver, W. W. and A. J. Kingston, "A Factor Analysis of the Cloze Procedure and Other Measures of Reading and Language Ability," The Journal of Communication. 13 (December, 1963) 252-261.